

Mathematical Preliminaries

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

<http://www.stanford.edu/~yyye>

Vectors and Norms

- Real numbers: \mathcal{R} , \mathcal{R}_+ , $\text{int } \mathcal{R}_+$
- n -dimensional Euclidean space \mathcal{R}^n , \mathcal{R}_+^n , $\text{int } \mathcal{R}_+^n$
- Component-wise: $\mathbf{x} \geq \mathbf{y}$ means $x_j \geq y_j$ for $j = 1, 2, \dots, n$
- $\mathbf{0}$: vector of all zeros; and \mathbf{e} : vector of all ones
- Inner-product of two vectors:

$$\mathbf{x} \bullet \mathbf{y} := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^n x_j y_j$$

- Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$,
Infinity-norm: $\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$,
 p -norm: $\|\mathbf{x}\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$

- The **dual** of the p norm, denoted by $\|\cdot\|^*$, is the q norm, where $\frac{1}{p} + \frac{1}{q} = 1$

- **Column vector:**

$$\mathbf{x} = (x_1; x_2; \dots; x_n)$$

Row vector:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

- **Transpose operation:** A^T
- A set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ is said to be **linearly dependent** if there are scalars $\lambda_1, \dots, \lambda_m$, not all zero, such that the **linear combination**

$$\sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0}$$

- A **linearly independent** set of vectors that span R^n is a **basis**.

Hyper plane and Half-spaces

$$H = \left\{ \mathbf{x} : \mathbf{a}\mathbf{x} = \sum_{j=1}^n a_j x_j = b \right\}$$

$$H^+ = \left\{ \mathbf{x} : \mathbf{a}\mathbf{x} = \sum_{j=1}^n a_j x_j \leq b \right\}$$

$$H^- = \left\{ \mathbf{x} : \mathbf{a}\mathbf{x} = \sum_{j=1}^n a_j x_j \geq b \right\}$$

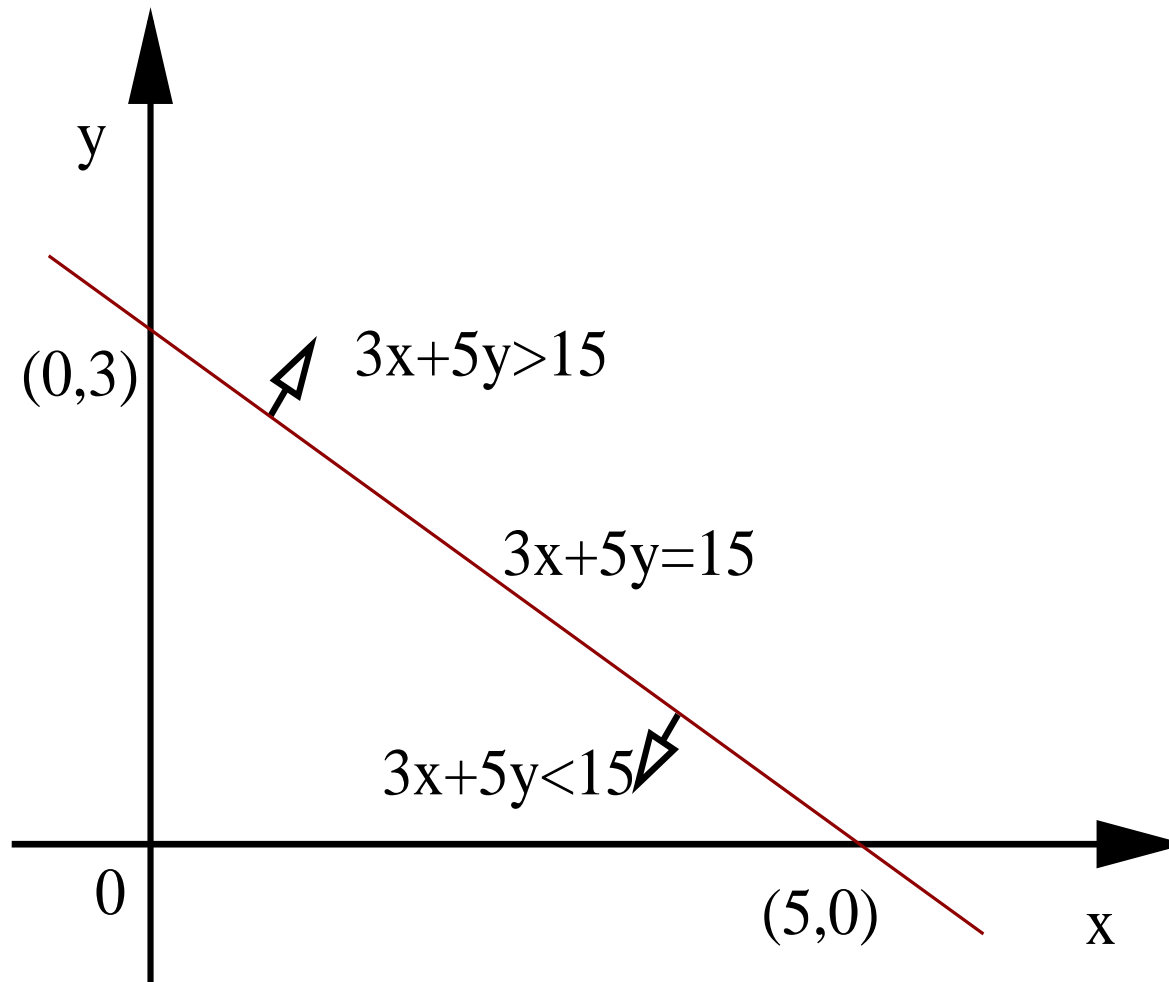


Figure 1: Plane and Half-Spaces

System of Linear Equations

Solve for $\mathbf{x} \in \mathcal{R}^n$ from:

$$\begin{array}{rcl} \mathbf{a}_1 \mathbf{x} & = & b_1 \\ \mathbf{a}_2 \mathbf{x} & = & b_2 \\ \dots & \cdot & \cdot \\ \mathbf{a}_m \mathbf{x} & = & b_m \end{array} \quad \Rightarrow \quad A\mathbf{x} = \mathbf{b}$$

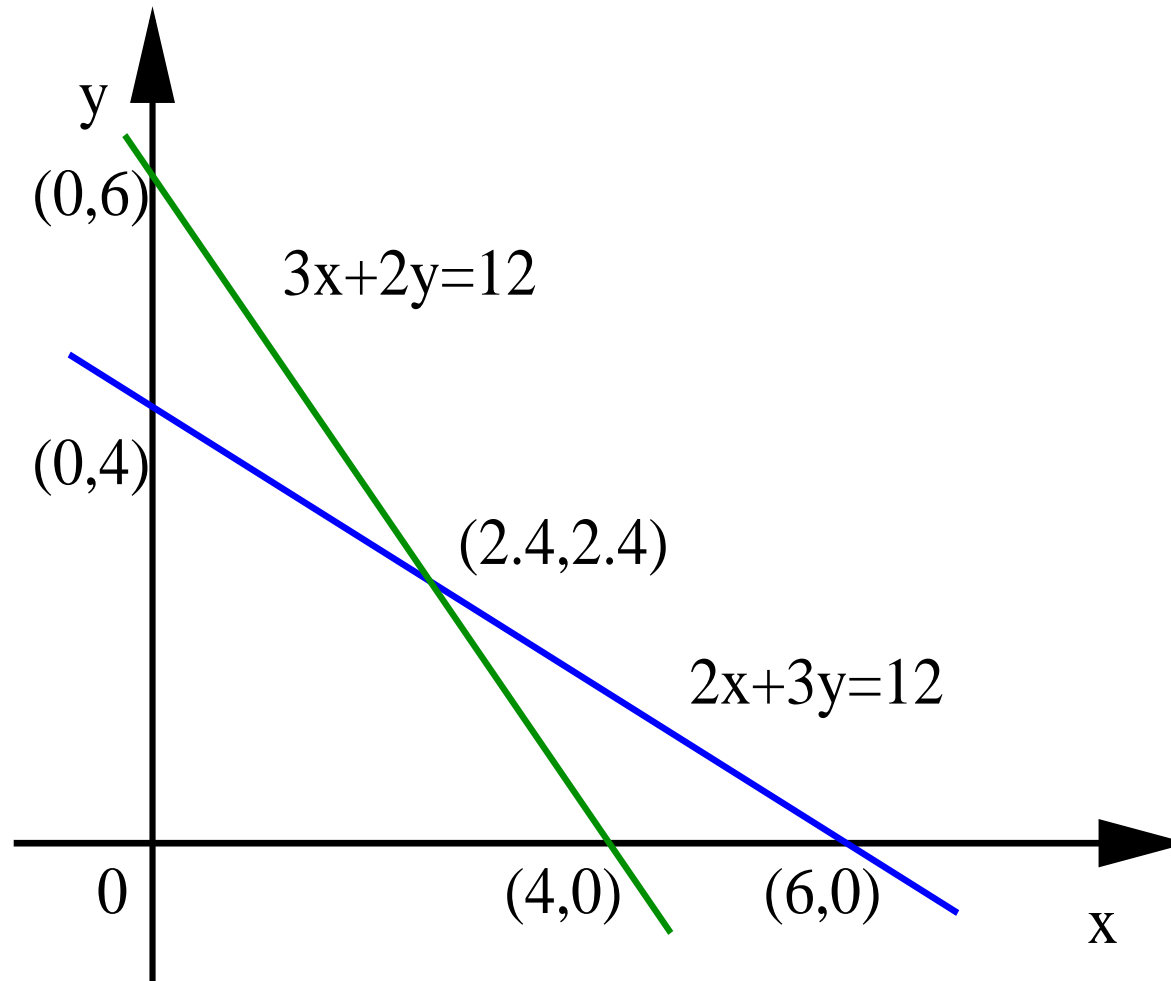


Figure 2: System of Linear Equations

Fundamental Theorem of Linear Equations

Theorem 1 Given $A \in \mathcal{R}^{m \times n}$ and $\mathbf{b} \in \mathcal{R}^m$, the system $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ has a solution if and only if that $A^T \mathbf{y} = \mathbf{0}$ and $\mathbf{b}^T \mathbf{y} \neq 0$ has no solution.

A vector \mathbf{y} , with $A^T \mathbf{y} = \mathbf{0}$ and $\mathbf{b}^T \mathbf{y} \neq 0$, is called an **infeasibility certificate** for the system.

Example Let $A = (1; -1)$ and $\mathbf{b} = (1; 1)$. Then, $\mathbf{y} = (1/2; 1/2)$ is an **infeasibility certificate**.

Alternative systems: $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ and $\{\mathbf{y} : A^T \mathbf{y} = \mathbf{0}, \mathbf{b}^T \mathbf{y} \neq 0\}$.

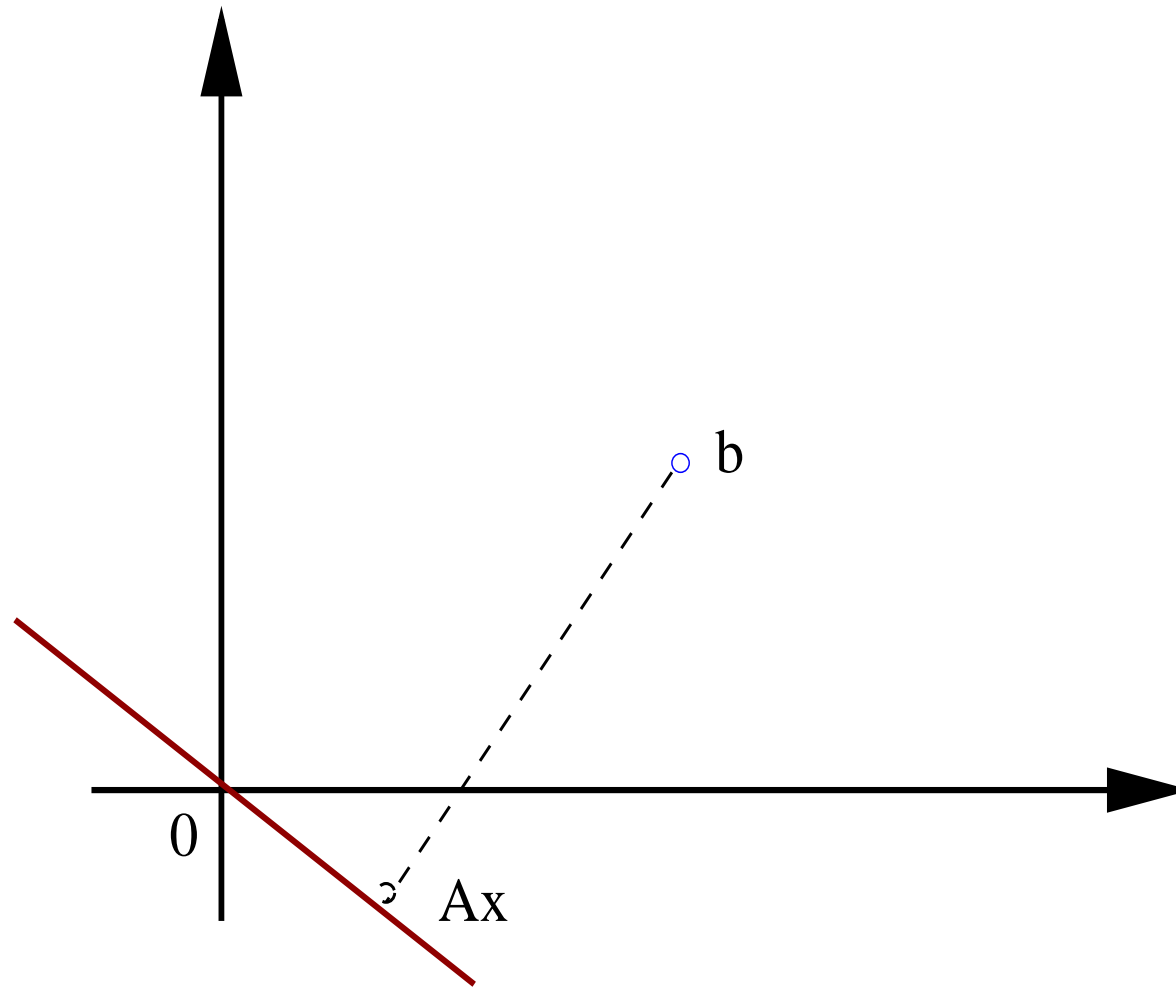


Figure 3: \mathbf{b} is not in the set $\{A\mathbf{x} : \mathbf{x}\}$, and \mathbf{y} is the distance vector from \mathbf{b} to the set.

Gaussian elimination method

$$\begin{pmatrix} a_{11} & A_{1.} \\ 0 & A' \end{pmatrix} \begin{pmatrix} x_1 \\ x' \end{pmatrix} = \begin{pmatrix} b_1 \\ b' \end{pmatrix}.$$

$$A = L \begin{pmatrix} U & C \\ 0 & 0 \end{pmatrix}$$

Matrices and Norms

- **Matrix:** $\mathcal{R}^{m \times n}$, i th row: $a_{i.}$, j th column: $a_{.j}$, ij th element: a_{ij}
- A_I denotes the **submatrix** of A whose rows belong to index set I , A_J denotes the **submatrix** whose columns belong to index set J , A_{IJ} denotes the **submatrix** whose rows belong to index set I and columns belong to index set J .
- **Determinant:** $\det(A)$, **Trace:** $\text{tr}(A)$
- **The operator norm of** $\|A\|$,

$$\|A\|^2 := \max_{0 \neq x \in \mathcal{R}^n} \frac{\|Ax\|^2}{\|x\|^2}$$

- **All-zero matrix:** $\mathbf{0}$, and **identity matrix:** I
- **Diagonal matrix:** $X = \text{diag}(\mathbf{x})$

- Symmetric matrix: $Q = Q^T$
- Positive Definite: $Q \succ 0$ iff $\mathbf{x}^T Q \mathbf{x} > 0$, for all $\mathbf{x} \neq \mathbf{0}$
- Positive Semidefinite: $Q \succeq 0$ iff $\mathbf{x}^T Q \mathbf{x} \geq 0$, for all \mathbf{x}
- Null space: $\mathcal{N}(A)$, Range space: $\mathcal{R}(A)$:

Theorem 2 *The null space and range space of a matrix are perpendicular to each other.*

Symmetric Matrix Space

- n -dimensional symmetric matrix space: \mathcal{M}^n

- Inner Product:

$$X \bullet Y = \text{tr} X^T Y = \sum_{i,j} X_{i,j} Y_{i,j}$$

- Frobenius norm:

$$\|X\|_f = \sqrt{\text{tr} X^T X}$$

- Positive semidefinite matrix set: \mathcal{M}_+^n , Positive definite matrix set: $\text{int } \mathcal{M}_+^n$

Affine and Convex Combination

$S \subset R^n$ is **affine** if

$$[\mathbf{x}, \mathbf{y} \in S \text{ and } \alpha \in R] \implies \alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in S.$$

When \mathbf{x} and \mathbf{y} are two distinct points in R^n and α runs over R ,

$$\{\mathbf{z} : \mathbf{z} = \alpha \mathbf{x} + (1 - \alpha)\mathbf{y}\}$$

is the **line** set determined by \mathbf{x} and \mathbf{y} .

When $0 \leq \alpha \leq 1$, it is called the **convex combination** of \mathbf{x} and \mathbf{y} and it is the **line segment** between \mathbf{x} and \mathbf{y} .

Convex Sets

- Set notations: $x \in \Omega$, $y \notin \Omega$, $S \cup T$, $S \cap T$
- Ω is said to be a **convex set** if for every $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$ and every real number $\alpha \in [0, 1]$, the point $\alpha\mathbf{x}^1 + (1 - \alpha)\mathbf{x}^2 \in \Omega$.
- The **convex hull** of a set Ω is the intersection of all convex sets containing Ω
- **Intersection** of convex sets is convex

Proof of convex set

- All solutions to the system of linear equations, $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$, form a convex set.
- All solutions to the system of linear inequalities,

$$\{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$$

, form a convex set.

- All solutions to the system of linear equations and inequalities, $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, form a convex set.
- **Ball** is a convex set: for $\mathbf{y} \in \mathcal{R}^n$ and $r > 0$,

$$B(\mathbf{y}, r) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| \leq r\}$$

- **Ellipsoid** is a convex set: for a positive definite matrix Q ,

$$E(\mathbf{y}, Q) = \{\mathbf{x} : (\mathbf{x} - \mathbf{y})^T Q (\mathbf{x} - \mathbf{y}) \leq 1\}$$

More proof of convex set

Consider the set B of all \mathbf{b} , for fixed A and \mathbf{c} , such that the following linear program is feasible:

$$\begin{aligned} & \text{minimize} && \mathbf{c}^T \mathbf{x} \\ & \text{subject to} && A\mathbf{x} = \mathbf{b}, \\ & && \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Show that B is a convex set.

Convex Cones

- A set C is a **cone** if $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C$ for all $\alpha > 0$
- A **convex cone** is cone plus convex-set.
- **Dual cone**:

$$C^* := \{\mathbf{y} : \mathbf{y} \bullet \mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in C\}$$

$-C^*$ is also called the **polar** of C .

Cone Examples

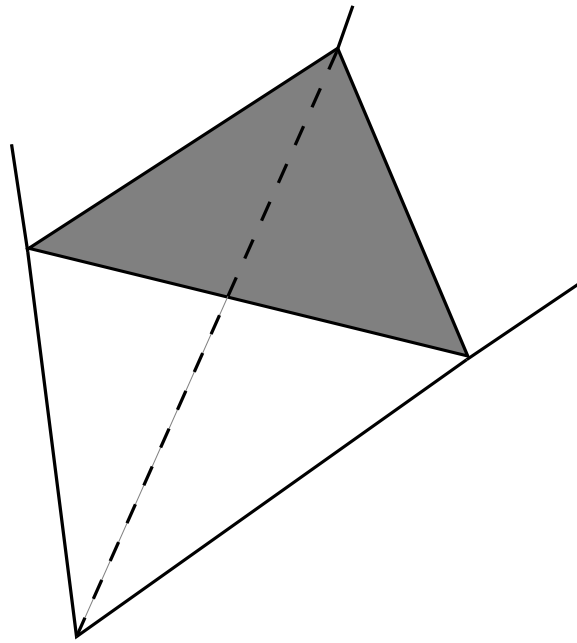
- Example 2.1: The n -dimensional non-negative orthant, $\mathcal{R}_+^n = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{x} \geq \mathbf{0}\}$, is a convex cone.
- Example 2.2: The set of all positive semi-definite matrices in \mathcal{M}^n , \mathcal{M}_+^n , is a convex cone, called the **positive semi-definite matrix cone**.
- Example 2.3: The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : t \geq \|\mathbf{x}\|\}$ is a convex cone in \mathcal{R}^{n+1} , called the **second-order cone**.
- Example 2.4: The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : t \geq \|\mathbf{x}\|_p\}$ is a convex cone in \mathcal{R}^{n+1} , called the **p -order cone**.

Polyhedral Convex Cones

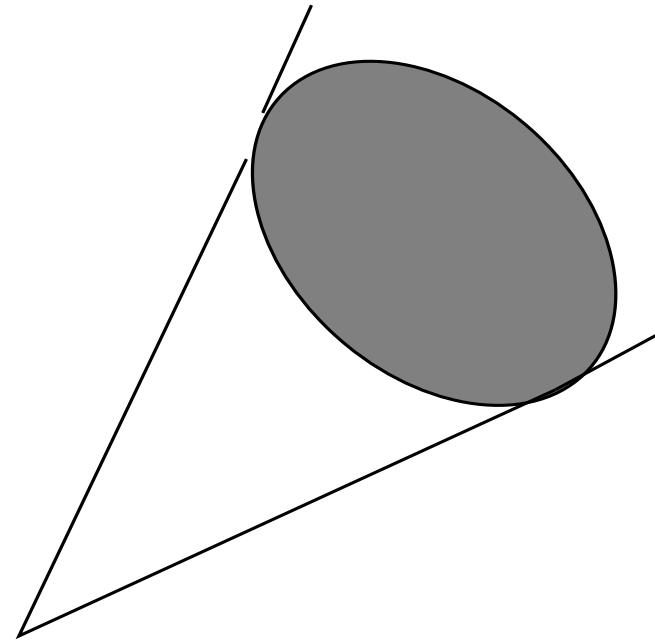
- A cone C is (convex) **polyhedral** if C can be represented by

$$C = \{\mathbf{x} : A\mathbf{x} \leq 0\}$$

for some matrix A .



Polyhedral Cone



Nonpolyhedral Cone

Figure 4: Polyhedral and non-polyhedral cones.

- The nonnegative orthant is a polyhedral cone but the second-order cone is not polyhedral.

Real Functions

- Continuous functions C
- Weierstrass theorem: a continuous function $f(\mathbf{x})$ defined on a compact set (bounded and closed) $\Omega \subset \mathcal{R}^n$ has a minimizer in Ω .

- The least upper bound or supremum of f over Ω

$$\sup\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

and the greatest lower bound or infimum of f over Ω

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}$$

- A function $f(\mathbf{x})$ is called homogeneous of degree k if $f(\alpha\mathbf{x}) = \alpha^k f(\mathbf{x})$ for all $\alpha \geq 0$.

Let $\mathbf{c} \in \mathcal{R}^n$ be given and $\mathbf{x} \in \text{int } \mathcal{R}_+^n$. Then $\mathbf{c}^T \mathbf{x}$ is homogeneous of

degree 1 and

$$\phi(\mathbf{x}) = n \log(\mathbf{c}^T \mathbf{x}) - \sum_{j=1}^n \log x_j$$

is homogeneous of degree 0.

Let $C \in \mathcal{M}^n$ be given and $X \in \text{int } \mathcal{M}_+^n$. Then $\mathbf{x}^T C \mathbf{x}$ is homogeneous of degree 2, $C \bullet X$ and $\det(X)$ are homogeneous of degree 1 and n , respectively; and

$$\Phi(X) = n \log(C \bullet X) - \log \det(X)$$

is homogeneous of degree 0.

- The gradient vector C^1 :

$$\nabla f(\mathbf{x}) = \{\partial f / \partial x_i\}, \quad \text{for } i = 1, \dots, n.$$

- The Hessian matrix C^2 :

$$\nabla^2 f(\mathbf{x}) = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} \quad \text{for } i = 1, \dots, n; j = 1, \dots, n.$$

- Vector function: $\mathbf{f} = (f_1; f_2; \dots; f_m)$
- The Jacobian matrix of \mathbf{f} is

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ \dots \\ \nabla f_m(\mathbf{x}) \end{pmatrix}.$$

Convex Functions

- f convex function iff for $0 \leq \alpha \leq 1$,

$$f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}).$$

- The level set of convex function f

$$L(z) = \{\mathbf{x} : f(\mathbf{x}) \leq z\}$$

is a convex set.

Proof of convex function

Consider the minimal-objective function of \mathbf{b} for fixed A and \mathbf{c} :

$$\begin{aligned} z(\mathbf{b}) &:= \text{minimize} && \mathbf{c}^T \mathbf{x} \\ &&& \text{subject to} && A\mathbf{x} = \mathbf{b}, \\ &&& && \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Show that $z(\mathbf{b})$ is a convex function in \mathbf{b} for all feasible \mathbf{b} .

Theorems on functions

Taylor's theorem or the mean-value theorem:

Theorem 3 Let $f \in C^1$ be in a region containing the line segment $[\mathbf{x}, \mathbf{y}]$. Then there is a α , $0 \leq \alpha \leq 1$, such that

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}).$$

Furthermore, if $f \in C^2$ then there is a α , $0 \leq \alpha \leq 1$, such that

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + (1/2)(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}).$$

Theorem 4 Let $f \in C^1$. Then f is convex over a convex set Ω if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$.

Theorem 5 *Let $f \in C^2$. Then f is convex over a convex set Ω if and only if the Hessian matrix of f is positive semi-definite throughout Ω .*

Known Inequalities

- **Cauchy-Schwarz**: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\mathbf{x}^T \mathbf{y} \leq \|\mathbf{x}\| \|\mathbf{y}\|$.
- **Arithmetic-geometric mean**: given $\mathbf{x} > \mathbf{0}$,

$$\frac{\sum x_j}{n} \geq \left(\prod x_j \right)^{1/n}.$$

- **Harmonic**: given $\mathbf{x} > \mathbf{0}$,

$$\left(\sum x_j \right) \left(\sum 1/x_j \right) \geq n^2.$$

Linear least-squares problem

Given $A \in \mathcal{R}^{m \times n}$ and $\mathbf{c} \in \mathcal{R}^n$,

$$(LS) \quad \begin{aligned} &\text{minimize} && \|A^T \mathbf{y} - \mathbf{c}\|^2 \\ &\text{subject to} && \mathbf{y} \in \mathcal{R}^m. \end{aligned}$$

$$AA^T \mathbf{y} = A\mathbf{c} \quad \text{or} \quad \mathbf{y} = (AA^T)^{-1} A\mathbf{c}$$

with the **projection**:

$$A^T \mathbf{y} = A^T (AA^T)^{-1} A\mathbf{c}$$

Projection matrix: $P = A^T (AA^T)^{-1} A$ or $P = I - A^T (AA^T)^{-1} A$

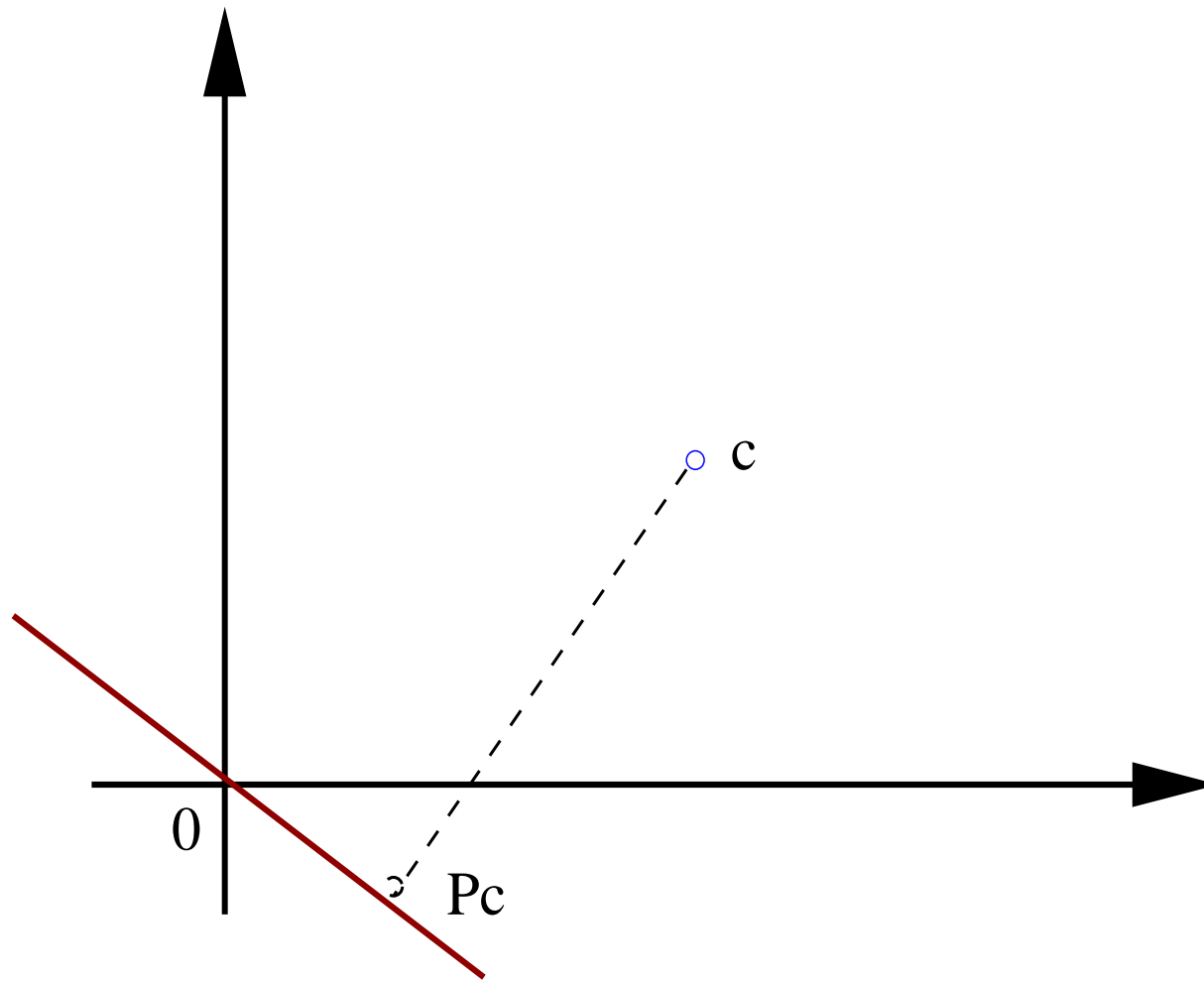


Figure 5: Projection of c onto a subspace

Choleski decomposition method

$$AA^T = L\Lambda L^T$$

$$L\Lambda L^T y^* = Ac$$

Solving ball-constrained linear problem

$$\begin{aligned} (BP) \quad & \text{minimize} \quad \mathbf{c}^T \mathbf{x} \\ & \text{subject to} \quad A\mathbf{x} = \mathbf{0}, \quad \|\mathbf{x}\|^2 \leq 1, \end{aligned}$$

\mathbf{x}^* minimizes (BP) if and only if there always exists a \mathbf{y} such that they satisfy

$$AA^T \mathbf{y} = A\mathbf{c},$$

and if $\mathbf{c} - A^T \mathbf{y} \neq \mathbf{0}$ then

$$\mathbf{x}^* = -(\mathbf{c} - A^T \mathbf{y}) / \|\mathbf{c} - A^T \mathbf{y}\|;$$

otherwise any feasible \mathbf{x} is a solution.

Solving ball-constrained linear problem

$$\begin{aligned} (BD) \quad & \text{minimize} \quad \mathbf{b}^T \mathbf{y} \\ & \text{subject to} \quad \|A^T \mathbf{y}\|^2 \leq 1. \end{aligned}$$

The solution \mathbf{y}^* for (BD) is given as follows: Solve

$$AA^T \bar{\mathbf{y}} = \mathbf{b}$$

and if $\bar{\mathbf{y}} \neq \mathbf{0}$ then set

$$\mathbf{y}^* = -\bar{\mathbf{y}} / \|A^T \bar{\mathbf{y}}\|;$$

otherwise any feasible \mathbf{y} is a solution.

System of nonlinear equations

Given $\mathbf{f}(\mathbf{x}) : \mathcal{R}^n \rightarrow \mathcal{R}^n$, the problem is to solve n equations for n unknowns:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}.$$

Given a point \mathbf{x}^k , **Newton's Method** sets

$$f(\mathbf{x}) \simeq f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) = \mathbf{0}.$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\nabla f(\mathbf{x}^k))^{-1} f(\mathbf{x}^k)$$

or solve for **direction vector** \mathbf{d}_x :

$$\nabla f(\mathbf{x}^k)\mathbf{d}_x = -f(\mathbf{x}^k) \quad \text{and} \quad \mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x.$$

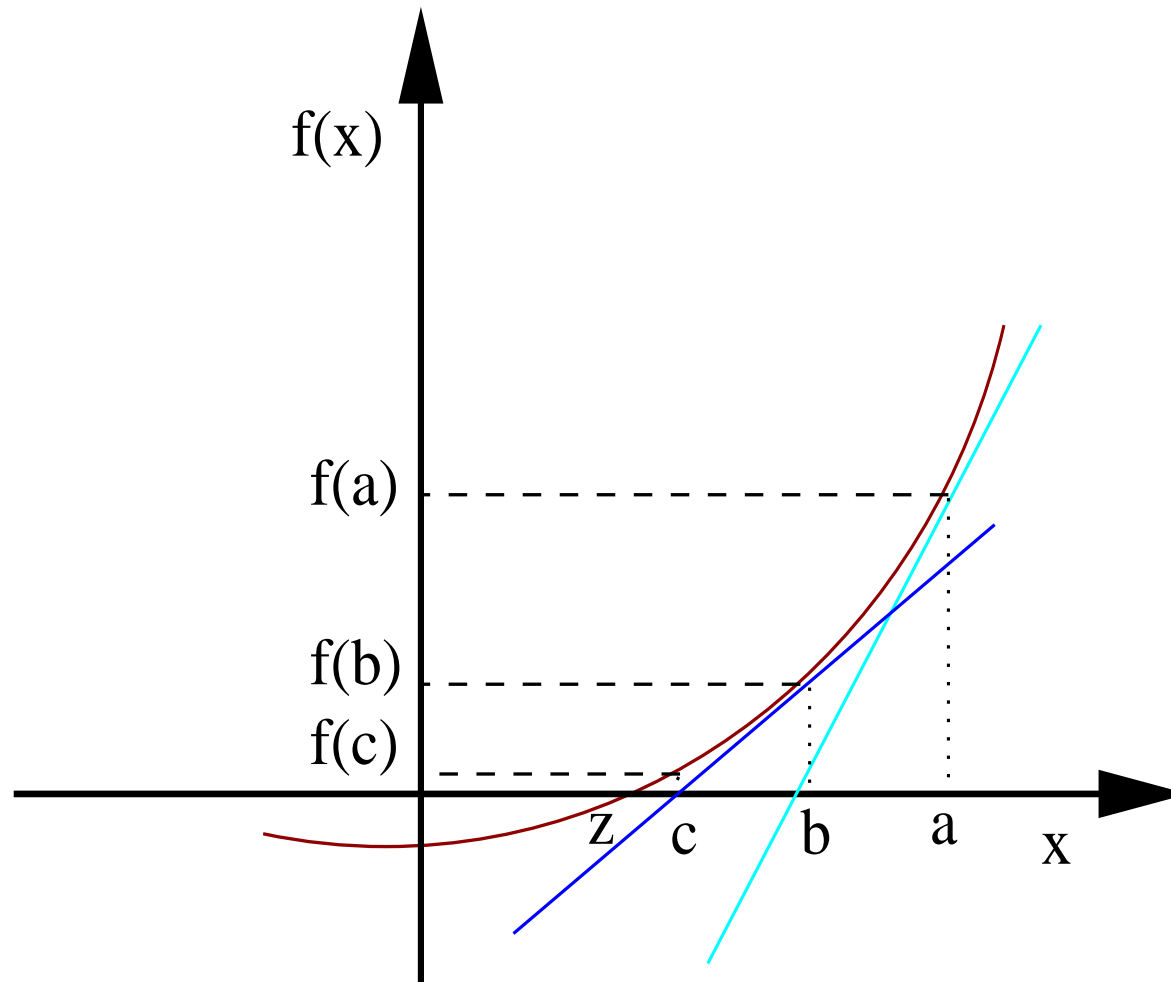


Figure 6: Newton's method for root finding

The quasi Newton method

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha (\nabla f(\mathbf{x}^k))^{-1} f(\mathbf{x}^k)$$

where scalar $\alpha \geq 0$ is called **step-size**. More generally

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha M^k f(\mathbf{x}^k)$$

where M^k is an $n \times n$ symmetric matrix. In particular, if $M^k = I$, the method is called the **gradient method**, where f is viewed as the gradient vector of a real function.

Convergence and Big O

- $\{\mathbf{x}^k\}_0^\infty$ denotes a sequence $\mathbf{x}^0, \mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k, \dots$
- $\mathbf{x}^k \rightarrow \bar{\mathbf{x}}$ iff

$$\|\mathbf{x}^k - \bar{\mathbf{x}}\| \rightarrow 0$$

- $g(x) \geq 0$ is a real valued function of a real nonnegative variable, the notation $g(x) = O(x)$ means that $g(x) \leq \bar{c}x$ for some constant \bar{c} ;
- $g(x) = \Omega(x)$ means that $g(x) \geq \underline{c}x$ for some constant \underline{c} ;
- $g(x) = \theta(x)$ means that $\underline{c}x \leq g(x) \leq \bar{c}x$.
- $g(x) = o(x)$ means that $g(x)$ goes to zero faster than x does:

$$\lim_{x \rightarrow 0} \frac{g(x)}{x} = 0$$